SUPPORTING INFORMATION

Raman spectra and structure of hydrogen-bonded water oligomers in tetrahydrofuran:H_2O binary solutions

Ankit Raj,* Yu-Jou Chen, and Chien-Lung Wang†
Department of Applied Chemistry and Institute of Molecular Science, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan

Hiro-o Hamaguchi‡
Department of Applied Chemistry and Institute of Molecular Science, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan and Center for Emergent Functional Matter Science, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan

---

Contents

* ankit@nycu.edu.tw
† kclwang@nycu.edu.tw
‡ hamaguchi@nycu.edu.tw

List of Figures

List of symbols

λ	: Wavelength	[nm]
$\delta\nu_s$	: Wavenumber spacing, vector	[cm^{-1}]
$\delta\nu_0$	: Wavenumber spacing at the fixed point, scalar	[cm^{-1}]

S1. Calibration of the Raman spectrometer

A. Wavenumber calibration

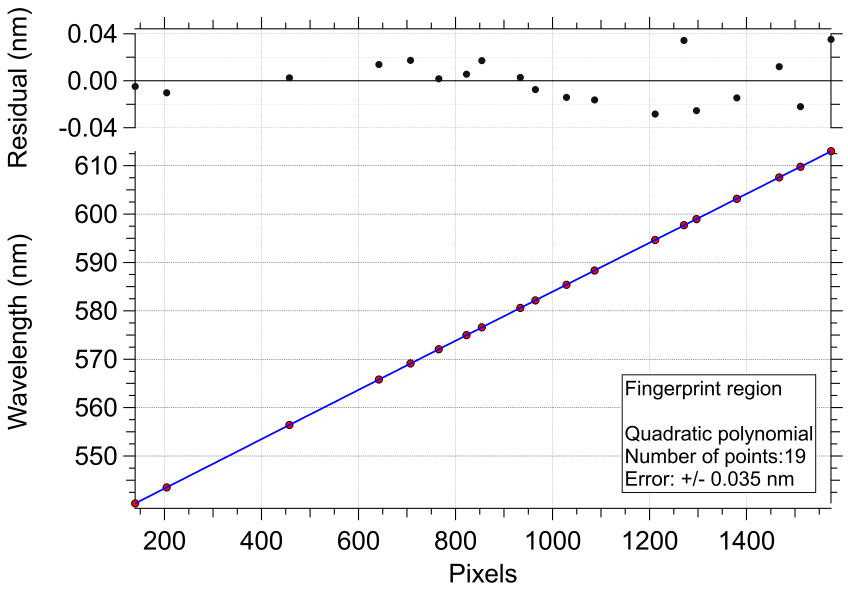


FIG. F1: Quadratic fit of positions of neon emission bands (in wavelength) to the corresponding observed band position (in pixels) for wavenumber calibration in the fingerprint region of the Raman spectra. A total of 19 data points were used. The maximal uncertainty encountered in the wavenumber calibration was 1 cm^{-1}.

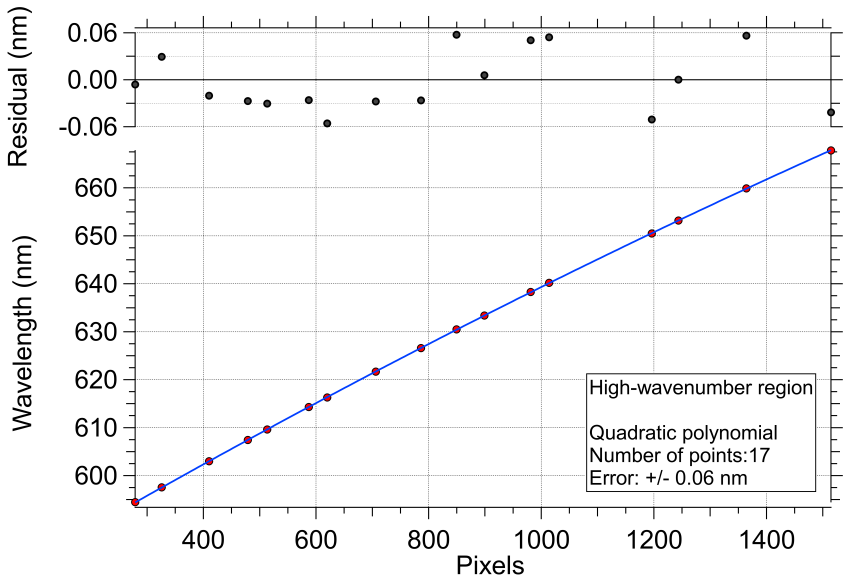


FIG. F2: Quadratic fit of positions of neon emission bands (in wavelength) to the corresponding observed band position (in pixels) for wavenumber calibration in the high-wavenumber region of the Raman spectra. A total of 17 data points were used. The eventual maximal uncertainty in the wavenumber calibration in the high-wavenumber region was 1.5 cm^{-1}.

B. Relative intensity calibration

We use a multi-step procedure for the intensity calibration of the Raman spectra which has been reported in our earlier work.[1, 2]

(i) The first correction affecting the Raman intensities is for the non-linear sampling of photons over the wavenumber scale. This correction is computed using the wavenumber step (or spacing) in the wavenumber array representing the x-axis, and is given by $C_0 = (\frac{\delta\nu_s}{\delta\nu_0})$. Here, $\delta\nu_s$ is the wavenumber step for every data-point of the 1D vector representing the wavenumber axis, and $\delta\nu_0$ is the similar wavenumber step at the laser frequency.

(ii) The next correction is for the channel-to-channel variations in the sensitivity of the detector. This correction is determined by fitting the spectrum from a broad-band white light source to the curve representing the number of photons per wavenumber emitted by a Black-body emitter, assuming certain temperature. The ratio of the broadband white-light spectrum to the obtained fit, gives us the second correction. This correction is defined as C_1.

The above described corrections were computed using the data analysis programs developed for IgorPro in our earlier study[2, 3], and were applied to the spectra covering the fingerprint and the high-wavenumber region, separately.

C. Correction for laser power variation

This correction is for correcting the acquired Raman spectral intensities which may vary due to temporal fluctuations in the incident laser power. The relative power of incident laser was monitored by a laser power meter in synchronization with Raman acquisition, in all the Raman measurements. By dividing the spectral intensity with the relative laser power, the intensity of all the THF-water mixtures in the Raman measurements were corrected. The average recorded laser power, in volts after ADC conversion, during each measurement (corresponding to each concentration) is shown in Fig. F3. Before correction of the spectral intensities, the recorded voltage values were normalized to unity.

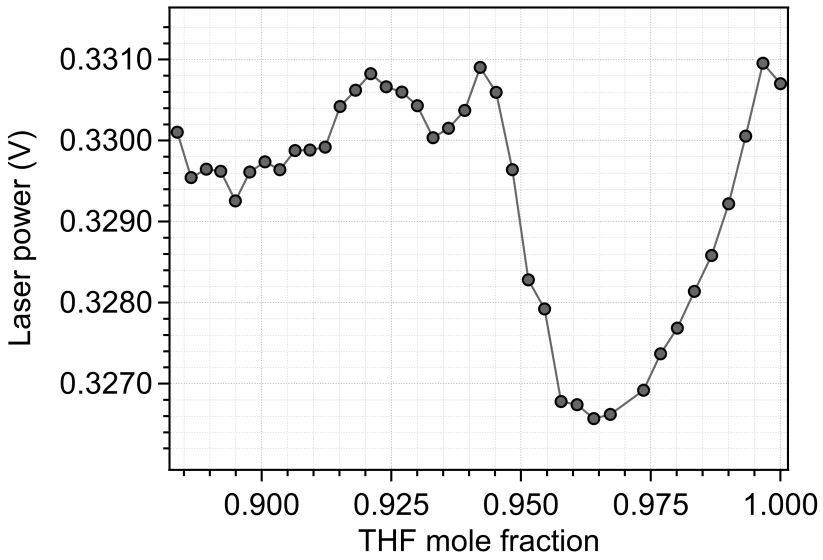


FIG. F3: The relative power of incident laser during all the Raman measurements.

D. Correction for density variation

The density of THF-water binary mixture changes depends on the concentrations of the components. This changes the number of molecules in the confocal volume when using a confocal Raman spectrometer for Raman measurements. In order to obtain the density correction curve, we fitted the known densities[4] and obtained the curve for the relative densities at the various THF-water concentrations relevant to the present

Raman experiments (shown in Fig. F4). By dividing the spectral intensity with the so obtained relative densities, the change to the number of the molecules in the confocal volume was corrected.

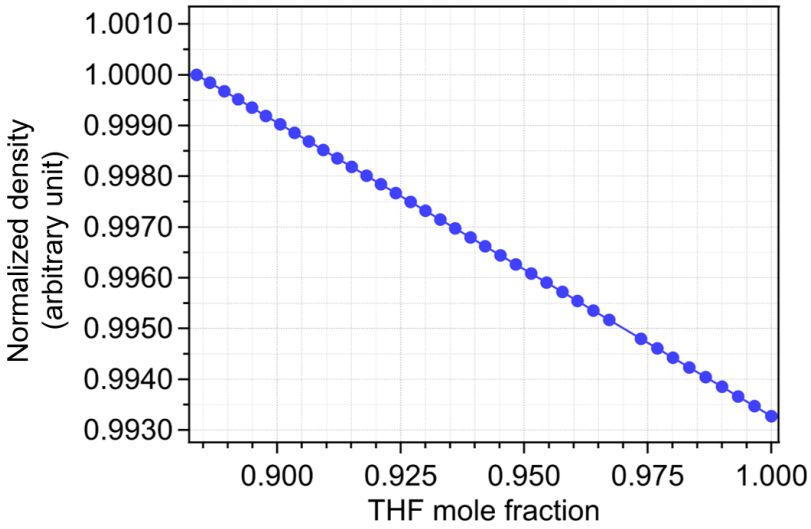


FIG. F4: Relative densities at the various concentrations of the THF:water binary solutions.

S2. Details on data analysis

A. Singular Value Decomposition (SVD)

Singular value decomposition (SVD) is the generalized eigen-decomposition of a real valued $m \times n$ matrix.[5, 6] It is a widely used tool to estimate the number of principal components dominating a two-dimensional spectral dataset. It is based on the theorem that any two dimensional matrix $\mathbb{A}_{(m,n)}$ can be expressed as a product of three matrices as shown below,

$$\mathbb{A} = \mathbb{U}\mathbb{W}\mathbb{V}^T \tag{E1}$$

where $\mathbb{W}_{(m,n)}$ is a diagonal matrix, while $\mathbb{U}_{(m,m)}$ and $\mathbb{V}_{(n,n)}$ are orthogonal matrices. From our experiments, a 2-D dataset of the Raman spectra corresponding to different concentrations are obtained. This dataset is the matrix $\mathbb{A}_{(m,n)}$. $\mathbb{U}_{(m,m)}$ and $\mathbb{V}_{(n,n)}$ correspond to the spectral and concentration-dependent components, respectively. The diagonal elements of the diagonal matrix $\mathbb{W}_{(m,n)}$ are the so-called singular values. See Fig. F5 for a pictorial representation.

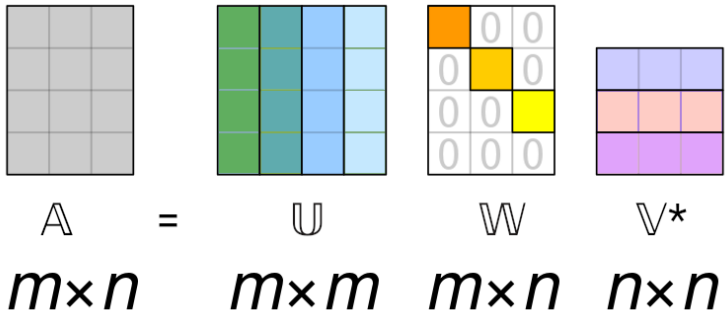


FIG. F5: Summary of matrix multiplication in SVD analysis.

For each such pair of column and rows in $\mathbb{U}_{(m,m)}$ and $\mathbb{V}_{(n,n)}$ matrices, respectively, there exists a singular value in $\mathbb{W}_{(m,n)}$. The relative magnitude of a singular value provides an estimate of the importance of the specific component in the whole dataset.

B. Hypothetical Addition Multivariate Analysis With Numerical Differentiation (HAMAND)

HAMAND (Hypothetical addition multivariate analysis with numerical differentiation)[7] combines the principle of the standard addition method with numerical chemometric method for separating and quantifying a known target spectral component from a complex spectrum. HAMAND processes a number of numerically generated hypothetical model spectra (S_j) which are obtained by adding the standard spectrum ($S_{\text{std.}}$) to the spectra of an unknown sample ($S_{\text{unkn.}}$) with different coefficients (c_j):

$$S_j = S_{\text{unkn.}} + c_j \times S_{\text{std.}} \tag{E2}$$

Subsequently, these hypothetical spectra are separated into two spectral components, the first W_1 (the spectrum corresponding to all the other substances except for the target), and W_2 (the target spectrum), with intensity profile of H_1 (as constant) and H_2 (which is linearly dependent on hypothetically added coefficient representing concentration, c_j). The calibration line which is obtained by plotting the intensity profile H_2 versus hypothetically added coefficient determines the amount of the target spectrum already contained in the sample.

C. Multi-Variate Curve Resolution Alternating Least Squares (MCR-ALS) analysis

MCR-ALS (multivariate curve resolution with alternating least-squares)[8] is a tool to separate a set of spectra, consisting of mixture of unknown spectral signatures, into several pure components by using multivariate statistical method with the alternating least squares. In the MCR procedure, the experimental data $\mathbb{A}_{(m,n)}$ is approximated by a linear combination of spectral components $\mathbb{W}_{(m,k)}$ and their intensity profiles $\mathbb{H}_{(k,n)}$ as,

$$\mathbb{A} \approx \mathbb{W}\mathbb{H} \tag{E3}$$

Matrix $\mathbb{A}_{(m,n)}$ is composed of n spectra with m data points in each Raman spectrum. The columns of matrix $\mathbb{W}$ represent the spectra of pure components, while the rows of matrix $\mathbb{H}$ represent their intensity profiles. The number of basic components, k, are typically guessed or estimated by singular value decomposition (SVD) analysis. MCR-ALS procedure relies on the iterative minimization of the Frobenius norm, $||\mathbb{A} - \mathbb{W}\mathbb{H}||^2$ while changing the elements of matrix $\mathbb{W}$ and $\mathbb{H}$. Non-negativity constraints ($\mathbb{W} \geq 0$ and $\mathbb{H} \geq 0$) are applied during the minimization, which results in convergence to physically meaningful spectra as the basic components, while also giving the intensity distribution.

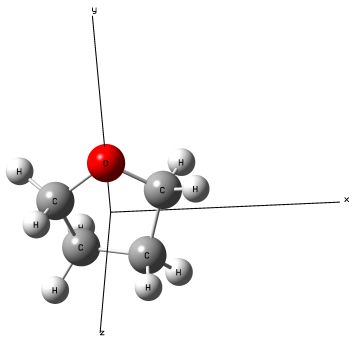
With regard to the present study, set of Raman spectra (having different concentration of THF:water) were supplied as the matrix $\mathbb{A}$. From the MCR-ALS analysis, sets of spectral components corresponding to pure THF, and THF:water complexes were obtained as $\mathbb{W}$. For the spectral components in $\mathbb{W}$, the corresponding concentration profiles were obtained as $\mathbb{H}$.

In Section 4.3 of the main document, we utilized the concentration profiles obtained form MCR-ALS analysis ($\mathbb{H}$) to determine the equilibrium constant for the formation of the 1:1 THF:water complex.

S3. Details on the quantum-chemical calculations

Optimized geometries of the studied molecules are shown using figures and described using atomic coordinates. The atomic coordinates are listed in "ATOM X Y Z" format, with distance in bohr.

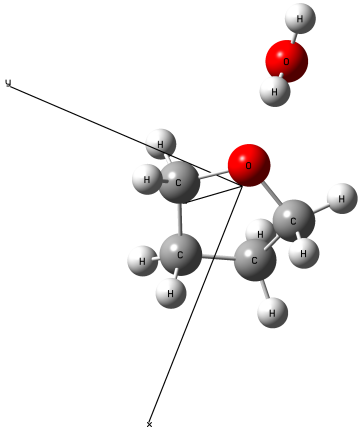
A. THF



Geometry (distances are in bohr)

C	-2.2059092718	0.8072720451	0.2490526572
O	0.0000075589	2.3624069296	-0.0000188973
C	2.2059224999	0.8072550375	-0.2490224216
C	1.3832567469	-1.8771328546	0.4288752028
C	-1.3832794236	-1.8771139573	-0.4288922104
H	-2.8894685153	0.9077903498	2.2010394479
H	-3.6752846115	1.5488874504	-0.9921892919
H	2.8895441043	0.9077903498	-2.2009846459
H	3.6752638245	1.5488477661	0.9922724399
H	2.5288539946	-3.3154373196	-0.4953538734
H	1.5002137781	-2.1717297979	2.4671563308
H	-1.5002421240	-2.1716693266	-2.4671771178
H	-2.5288898994	-3.3154203120	0.4953141891

B. THF + 1 water

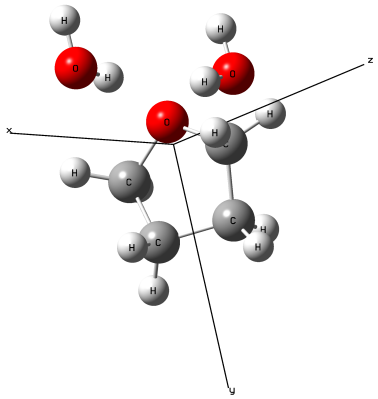


Geometry (distances are in bohr)

C	0.6158635894	-2.0429638690	-1.2015728212
O	-0.5557892002	0.3706603040	-1.6556853146
C	0.8604092501	2.3320220254	-0.3832496586
C	3.1040261143	1.0256279933	0.8971927665
C	2.1552343797	-1.6955377460	1.2046511848
H	-0.8806746716	-3.4461855710	-1.0401391992
H	1.8252428687	-2.5301900309	-2.8089775265
H	1.4337445562	3.7399359170	-1.7748589944
H	-0.3939171618	3.2262743782	0.9897931194
H	4.7619091803	1.0673607020	-0.3282454042
H	3.6079309978	1.9173273264	2.6810940997

```
H         3.6751617793      -3.0708897691       1.3846513647
H         0.9268595748      -1.8544391352       2.8522805976
O        -5.0797837683      -0.0254016967       1.0329336740
H        -6.5940495488       0.2770565066       0.0793212484
H        -3.6908238283       0.1967828364      -0.1550368996
```

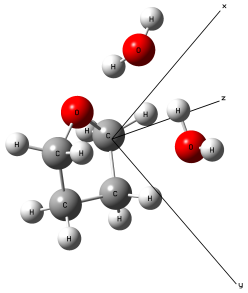
C. THF + 2 water molecules (with second water hydrogen bonded to THF)



Geometry (distances are in bohr)

```
C  -0.2116228546       0.6928359084       2.2385089348
O   0.0505350524      -0.8718798869      -0.0055557944
C   0.1748733533       0.7254998221      -2.2397013519
C   0.5325380117       3.4090278889      -1.2578356331
C  -0.8739283499       3.3224765489       1.2735619327
H  -1.6679704233      -0.1486968689       3.4228606831
H   1.5894126242       0.6749288649       3.2462998045
H   1.7311723090       0.0550779537      -3.4062802273
H  -1.5988442466       0.5148709634      -3.2730847807
H   2.5340658589       3.7988840295      -0.9563090646
H  -0.2158123771       4.8185102531      -2.5557221187
H  -0.2900238064       4.8015178370       2.5784914272
H  -2.9071809170       3.4815971463       0.9744050807
O  -4.7363355185      -3.0410264293      -0.2275419063
H  -4.8380367917      -4.8451799560      -0.3962868782
H  -2.9423713944      -2.6373374945      -0.1960420638
O   4.9941716221      -2.6583569166       0.2211697503
H   3.1746999766      -2.3932396988       0.1832221627
H   5.2327570870      -4.4498643969       0.3866568345
```

D. THF + 2 water molecules (with second water hydrogen bonded to the first water molecule)



Geometry (distances are in bohr)

C	-1.4649042480	-0.9365387513	2.2867234037
O	-0.5265513597	-2.2056144746	0.0533601927
C	-1.4410672444	-0.9284771802	-2.1871102777
C	-3.6056538780	0.7462131086	-1.2931224865
C	-2.7233502626	1.4897465858	1.3649604199
H	-2.8039546349	-2.1987358720	3.2256488789
H	0.1239754735	-0.5857470263	3.5508367065
H	0.0941593768	0.2035064814	-2.9765659861
H	-2.0026843595	-2.3693573417	-3.5456248422
H	-3.8949690371	2.3662506322	-2.5277239385
H	-5.3639078501	-0.3297779720	-1.2075859293
H	-1.3288137412	2.9981107521	1.2501898017
H	-4.2625056241	2.0840049380	2.5947222837
O	3.2092366087	3.3169679976	-0.0974153747
H	3.9929267632	1.6482681401	-0.0799221812
H	4.4560400215	4.4333103973	-0.7992180329
O	4.5855127079	-1.7985221432	-0.1602582125
H	5.4588968166	-2.7070060225	1.1454479593
H	2.7951088275	-2.2711463924	-0.0243869139

Bibliography

[1] A. Raj, C. Kato, H. A. Witek, and H. Hamaguchi, Toward standardization of raman spectroscopy: Accurate wavenumber and intensity calibration using rotational Raman spectra of H_2, HD, D_2, and vibration–rotation spectrum of O_2, Journal of Raman Spectroscopy **51**, 2066 (2020).

[2] A. Raj, C. Kato, H. A. Witek, and H. Hamaguchi, Accurate intensity calibration of multichannel spectrometers using Raman intensity ratios, Journal of Raman Spectroscopy (2021).

[3] A repository containing procedures for Raman data analysis in IgorPro., `https://github.com/ankit7540/RamanSpec_BasicOperations` (2022), accessed: 2022-02-25.

[4] O. Kiyohara, P. J. D'Arcy, and G. C. Benson, Thermal expansivities of water tetrahydrofuran mixtures at 298.15 K, Canadian Journal of Chemistry **56**, 2803 (1978).

[5] J. W. Demmel, *Applied Numerical Linear Algebra* (Society for Industrial and Applied Mathematics, 1997).

[6] G. H. Golub and C. F. van Loan, *Matrix Computations*, 4th ed. (JHU Press, 2013).

[7] M. Ando, I. K. Lednev, and H. Hamaguchi, Quantitative spectrometry of complex molecular systems by hypothetical addition multivariate analysis with numerical differentiation (HAMAND), in *Frontiers and Advances in Molecular Spectroscopy* (Elsevier, 2018) pp. 369–378.

[8] M. Ando and H. Hamaguchi, Molecular component distribution imaging of living cells by multivariate curve resolution analysis of space-resolved Raman spectra, Journal of Biomedical Optics **19**, 011016 (2013).
